

تحلیل توزیع مکانی و زمانی رخداد‌های طبیعی ایران از سال ۱۳۹۰ تا ۱۴۰۰ بر پایه گزارش‌های خبری

مهدی شاخصی^{۱*}، علی اصغر آل شیخ^۲

^۱ کارشناس ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین

طوسی

m_shakhesi@email.kntu.ac.ir

^۲ استاد دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی

alesheikh@kntu.ac.ir

(تاریخ دریافت تیر ۱۴۰۱، تاریخ تصویب شهریور ۱۴۰۱)

چکیده

رخداد‌های طبیعی اغلب با تأثیرات اجتماعی، اقتصادی و محیطی همراه می‌شوند. هر چند سنجنده‌های مختلفی در امر پایش این رخداد‌ها می‌توانند بکار گرفته شوند ولی تحلیل کارآمد این رخداد‌ها نیازمند توسعه پایگاه‌داده‌هایی متشکل از رخداد‌ها و اطلاعات مکانی، زمانی و توصیفی آن‌ها است. برای توسعه سیاهه رخداد‌ها، می‌توان از رسانه‌های جمعی و اجتماعی به خاطر دربرگرفتن اطلاعات جزئی‌تر در مقایسه با گزارش‌های سالیانه، بهره برد. در این پژوهش، با کمک تکنیک‌های متن‌کاوی چارچوبی برای استخراج خودکار رخداد‌های طبیعی از گزارش‌های خبری توسعه داده شده است. پس از وب‌کاوی اخبار و حذف موارد غیرمرتبط با رخداد‌های طبیعی، استانداردسازی متن انجام می‌شود. در ادامه، با کمک مدل‌های مبتنی بر یادگیری ماشین نخست مشخص می‌شود که خبر درباره رخدادی جدید یا جاری است و دوم در قالب چه موضوعی به رخداد طبیعی پرداخته شده است (مقادیر ۰/۸۷۵ و ۰/۷۱۶ به ترتیب برای F-score مدل‌های طبقه‌بندی اول و دوم اخذ شده است). همچنین با بهره‌گیری از مدلی قاعده محور انواع رخداد‌های مطرح‌شده شناسایی می‌شود. بعد از تشخیص جاینام‌های مدیریتی، به ازای هر «نوع رخداد-جاینام» اطلاعات زمانی و توصیفی از هر خبر استخراج می‌شود. در آخر، رخداد‌های طبیعی به همراه گزارش‌های خبری مربوطه گردآوری می‌شوند. نتایج پیاده‌سازی این چارچوب برای کشور ایران از آغاز ۱۳۹۰ تا پایان ۱۳۹۹ براساس اخبار خبرگزاری مهر در بخش‌های بررسی محتوای خبری مرتبط با رخداد‌های طبیعی، تحلیل توزیع مکانی و زمانی رخداد‌ها ارائه شده است. طبق نتایج تحقیق، از نظر زمانی تعداد کل رخداد‌ها در ماه‌های آغازین و پایانی سال بیشتر است. از لحاظ مکانی، فارغ از حجم خسارات، از استان‌های جنوبی بیشتر از استان‌های شمالی رخداد زمین‌لرزه گزارش می‌شود. در مقابل، تنوع رخداد‌های هواشناسی در استان‌های شمالی بیشتر است. خروجی این تحقیق می‌تواند در افزایش کارآمدی مدیریت بحران مورد استفاده قرار گیرد.

واژگان کلیدی: رخداد طبیعی، توزیع مکانی و زمانی، گزارش‌های خبری، متن‌کاوی، استخراج اطلاعات

* نویسنده رابط

۱- مقدمه

رخدادهای طبیعی همواره در تاریخ بشر بوده اند اما با افزایش جمعیت و تغییرات اقلیمی، میزان تلفات همواره روندی افزایشی داشته است. آنچه که در مرحله اول برای مدیریت بهتر این رخدادها و کاهش خسارات تبیین می‌شود، ثبت رخدادها می‌باشد. در ثبت حوادث طبیعی، سنجنده‌های گوناگونی می‌توانند مشارکت کنند که خروجی این سنجنده‌ها می‌تواند تصاویر سنجش‌ازدوری، داده‌های زمین‌شناسی، اطلاعات هواشناسی و دانش شهروندی باشد [۱، ۲]. صرف نظر از اینکه برای یک نوع حادثه طبیعی مشخص کدام منابع داده کارآمد هستند، باید به شش پرسش چه زمانی، چه مکانی، چه کسی، چه چیزی، چرا و چگونه به ازای هر رخداد طبیعی پاسخ داده شود [۳]. هرچند داده تهیه‌شده از این منابع بصورت موردی مورد توجه قرار می‌گیرد ولی آنچه که معمولاً به عنوان خروجی نهایی منتشر می‌شود، گزارش‌های سالیانه و اطلاعات کلی از مخاطرات طبیعی است. برای بررسی اثر این رخدادها و کشف روابط احتمالی بین آن‌ها، ضروری است که پایگاه داده‌ای برای رخدادهای طبیعی تهیه شود [۴].

شبکه‌های اجتماعی و رسانه‌های جمعی نظیر رادیو، تلویزیون و روزنامه‌ها با توجه به اهمیت‌شان در اطلاع‌رسانی به مردم، جزء منابع مهم برای گردآوری رخدادهای طبیعی شناخته می‌شوند. از این دو گروه، شبکه‌های اجتماعی معمولاً محتوای با قدرت تفکیک زمانی و مکانی بالاتری را ارائه می‌کنند [۵]. از طرفی، فارغ از جهت‌گیری‌هایی که رسانه‌های جمعی اتخاذ می‌کنند و در عین تلاشی که برای جذب مخاطب بیشتر دارند، اغلب موفق بودن گزارشات خبری را نیز در نظر می‌گیرند [۶-۹]. از این رو، سهم اطلاعات رسمی منتشرشده در رسانه‌های جمعی در مقایسه با شبکه‌های اجتماعی بیشتر است. بطور ویژه در زمینه رخدادهای طبیعی، این دو نوع رسانه در مقاطع زمانی مختلف پس از وقوع رخداد، موضوعات متفاوتی را مورد پوشش قرار می‌دهند. در حالیکه معمولاً فرکانس انتشار محتوا در شبکه‌های اجتماعی درباره یک رخداد طبیعی پس از گذشت چند هفته با یک روند نزولی همراه می‌شود؛ در رسانه‌های جمعی، تحت تاثیر رویدادهای سیاسی یا مدیریتی نظیر نشست‌های خبری و بازدید مسئولین ذیربط از منطقه، گزارش‌ها در یک روند موجی شکل منتشر می‌شوند [۱۰].

در تهیه سیاهه رخدادهای طبیعی، از نقطه نظر نحوه استخراج رخدادها از منابع متنی، استخراج دستی روش غالب در تحقیقات پیشین می‌باشد. دلیل این امر آن است که بیشتر پژوهش‌ها تنها در حوزه بررسی توزیع مکانی و زمانی رخدادها تعریف شده است. این چالش در مطالعاتی که روی چندین نوع رخداد طبیعی انجام شده است، به دلیل پیچیدگی‌های موجود در اختصاص نوع رخداد به محتوای متنی (در قالب طبقه بندی متن) و شناسایی موضوع شدت می‌یابد. همچنین تفاوت‌هایی که در نحوه پوشش خبری انواع مخاطرات طبیعی هست، مرحله ردیابی موضوع را با دشواری مواجه می‌سازد. بطور ویژه در استخراج انواع مختلف رخدادهای طبیعی از گزارش‌های خبری، همچنان که ممکن است چندین گزارش درباره رخدادی واحد منتشر شده باشد، در یک گزارش خبری نیز چندین رخداد می‌تواند ذکر شود. پژوهش حاضر در تلاش است با ارائه چارچوبی برای استخراج خودکار رخدادهای طبیعی این چالش‌ها را حل کند. نتیجه پیاده‌سازی این چارچوب برای گردآوری مخاطرات طبیعی ایران (از آغاز ۱۳۹۰ تا پایان ۱۳۹۹) با تحلیل توزیع مکانی و زمانی این حوادث ارائه می‌شود.

در ادامه، در بخش ۲، تحقیقات انجام‌گرفته در زمینه استخراج رخدادهای طبیعی از متون غیرساختاریافته بررسی می‌شوند. در بخش ۳، چارچوب پیشنهادی ارائه شده و مراحل به تفصیل بیان می‌شود. پس از ارزیابی، نتایج پژوهش در بخش ۴ ارائه‌شده و در بخش ۵ به نتیجه‌گیری تحقیق پرداخته می‌شود.

۲- پیشینه تحقیق

با توجه به هدف تعریف‌شده، استخراج رخدادهای طبیعی از متون خبری یا محتوای به‌اشتراک‌گذاشته‌شده در شبکه‌های اجتماعی می‌تواند بصورت بیدرنگ یا غیربیدرنگ انجام شود. در ادامه، با مروری بر تحقیقات انجام شده، این دو حالت مورد بررسی اجمالی قرار می‌گیرد.

حالت بیدرنگ برای زمانی کاربرد دارد که پایش یک یا چند رخداد طبیعی جاری برای سامانه مدیریت بحران مدنظر باشد. در این حالت، شبکه‌های اجتماعی از این نظر که بخش عظیمی از مطالب توسط خود مردم منتشر می‌شود و زمینه تسریع فرآیند امداد رسانی را فراهم

می‌سازند، کاربرد بیشتری دارند. برای نمونه، Bae و Choi، سامانه پایش بیدرنگی را برای مدیریت بحران توسعه داده اند. این سامانه ابتدا محتوای به اشتراک گذاشته شده به زبان کره‌ای تویتر را گردآوری می‌کند و سپس پیش‌پردازش‌های زبانی انجام شده و محتوای مربوط به موارد بحرانی براساس کلیدواژه‌های تعریف شده شناسایی می‌شود؛ محتوا در یکی از ۷۱ نوع رخداد از جمله رخدادهای طبیعی طبقه‌بندی می‌شود. در آخر، براساس فرکانس محتوای طبقه‌بندی شده، مناطقی که بیشترین ارجاع را داشته‌اند و نیز کلیدواژه‌های پر استفاده در سامانه نمایش داده می‌شود [۱۱]. برخی مطالعات متن‌کاوی شبکه‌های اجتماعی تنها به مقطعی از فرآیند کلی محدود شده اند و به موضوعاتی مثل تشخیص وقوع زلزله [۱۲] و برآوردی از بزرگی زلزله [۱۳]، افزایش آگاهی موقعیتی^۱ با ارائه راهکارهایی برای حل چالش‌های این حوزه نظیر طبقه‌بندی متن، تجزیه مکانی^۲ و محاسبه اهمیت رخداد [۱۴-۱۸]، بررسی خسارات حادثه با استفاده از عقیده‌کاوی^۳ [۱۹-۲۲] و بهره‌بردن از بخش تصویری محتوا در کنار بخش متنی [۲۳-۲۵] پرداخته‌اند.

در حالت غیربیدرنگ، گردآوری رخدادهای طبیعی مدت‌ها پس از وقوع این حوادث انجام می‌شود. گروهی از پژوهش‌ها به لحاظ زمانی تنها حول رخدادی خاص که معمولاً بازتاب خبری گسترده‌ای داشته است، معطوف می‌شوند. در این گونه مطالعات، هدف آن است که تغییرات انتشار محتوای شبکه‌های اجتماعی از قبل وقوع حادثه تا مدتی پس از آن مورد بررسی قرار گیرد تا براساس روندهای اتفاق افتاده، برای حوادث بعدی از آن نوع رخداد، آمادگی‌های لازم فراهم شود [۲۶-۲۹]. گروه دیگری از تحقیقات با هدف تهیه سیاهه‌ای از اطلاعات رخدادهای طبیعی، حوادث چندساله را گردآوری می‌کنند که می‌تواند تنها برای یک نوع رخداد طبیعی مشخص باشد یا تمامی رخدادهای محتمل در منطقه مورد مطالعه را پوشش دهد. در این دسته تحقیقات نیز استفاده از رسانه‌های جمعی بویژه روزنامه‌ها و خبرگزاری‌های برخط با توجه به رسمی بودن گزارش‌های خبری، بیشترین کاربرد را دارند.

Franceschini و همکاران با هدف ساخت پایگاه‌داده‌ای برای زمین‌لغزش، چارچوبی را توسعه داده اند که اخبار سال‌های ۲۰۱۰ تا ۲۰۱۹ به کمک سرویس اخبار گوگل گردآوری شده و موارد مرتبط با زمین‌لغزش شناسایی می‌شود. در این پژوهش، موقعیت رخدادهای در سطح شهرستان لحاظ شده و بر این اساس، توزیع مکانی زمین‌لغزش در ایتالیا مورد بررسی قرار گرفته است [۳۰]. Badoux و همکاران، با تمرکز بر تلفات جانی ناشی از رخدادهای طبیعی، با استفاده از متن کاوی اخبار، توزیع مکانی و زمانی مخاطرات طبیعی را برای کشور سوئیس و در بازه ۱۹۴۶ تا ۲۰۱۵ مورد مطالعه قرار داده اند [۳۱]. در این تحقیق، در کنار پایگاه‌داده‌های موجود درباره سیل، زمین‌لغزش و بهمن، از روزنامه Neue Zürcher Zeitung (NZZ) به عنوان روزنامه‌ای ملی برای ساخت پایگاه داده جامع جدید استفاده شده است. Kirschbaum و همکاران با استفاده از گزارش‌های خبری، مقالات علمی و سایر پایگاه داده‌های موجود، سیاهه‌ای را در مقیاس جهانی توسعه داده‌اند (محدود به سال‌های ۲۰۰۳، ۲۰۰۷ و ۲۰۰۸) که تنها شامل زمین‌لغزش‌هایی است که بر اثر بارش باران اتفاق افتاده اند [۳۲]. در پژوهش دیگری، Froude و همکاران پایگاه داده جهانی زمین‌لغزش‌های غیرلرزه‌ای را با تمرکز بر حجم تلفات جانی برای سال‌های ۲۰۰۴ تا ۲۰۱۶ ارائه کرده اند که اطلاعات اولیه از رسانه‌های جمعی گردآوری شده است [۳۳]. Han و همکاران با رویکردی مقایسه‌ای، اطلاعات استخراج شده از اخبار و اطلاعات اعلام شده در گزارش‌های رسمی سازمان‌های ذیربط را درباره زمین‌لرزه، برای کشور چین و سال‌های ۲۰۱۵ تا ۲۰۱۷ از لحاظ کمی و توزیع مکانی و زمانی مورد بررسی قرار داده اند [۳۴]. Zezere و همکاران، با استفاده از گزارش‌های خبری ۱۶ روزنامه ملی و منطقه‌ای، پایگاه‌داده رخدادهای طبیعی هیدروژئومورفولوژیکی (سیل و زمین‌لغزش) را برای کشور پرتغال و بازه زمانی از سال ۱۸۶۵ تا ۲۰۱۰ با محوریت خسارات جانی توسعه داده اند [۳۵]. در برخی مطالعات، علاوه بر اخبار از علم شهروندی^۴ نیز بهره گرفته شده است که از آن جمله می‌توان به تحقیقات Valenzuela و همکاران [۳۶]، Juang و همکاران [۳۷] اشاره کرد. همچنین در مقیاس جهانی و برای رخداد سیل، de Bruijn

^۱ situation awareness

^۲ geoparsing

^۳ sentiment analysis

^۴ citizen science

و همکاران با استفاده از محتوای تویتر پایگاه داده‌ای را توسعه داده اند که در وبگاه globalfloodmonitor.org قابل دسترس است [۳۸]. برخلاف حالت بیدرنگ که استخراج خودکار رخدادهای طبیعی را می‌طلبد، آنچه که در پژوهش‌های متمرکز بر حالت غیربیدرنگ و بویژه تهیه پایگاه داده رخدادهای طبیعی اتخاذ شده است، روش غیرخودکار است. برای نمونه، در تحقیق Liu و همکاران با موضوع تحلیل مکانی و زمانی توزیع رخدادهای طبیعی در چین، هرچند از طبقه‌بندی متن برای شناسایی رخداد مرتبط با محتوای خبر استفاده شده است ولی استخراج رخدادها بصورت دستی انجام گرفته است [۴].

۳- روش تحقیق

همچنان که در شکل ۱ نشان داده شده است، فرآیند استخراج رخدادهای طبیعی با وبکاوی شروع می‌شود؛ گزارش‌های خبری ذخیره‌شده مورد بررسی قرار می‌گیرد تا موارد غیرمرتبط حذف شود؛ متن اخبار از لحاظ نوع نگارش و واژگان استانداردسازی می‌شود و نیز محتوای متنی به عنوان خبر و سه جمله آغازین متن اصلی محدود می‌شود؛ طبقه‌بندی گزارش‌های خبری انجام می‌شود؛ در ادامه، جاینام‌های^۱ مدیریتی شامل نام استان‌ها، شهرستان‌ها و شهرها شناسایی می‌شود؛ در قالب یک خوشه‌بندی مکانی و زمانی، به ازای هر مورد رخداد، اولین گزارش خبری رخداد شناسایی شده و گزارش‌های بعدی درباره آن گردآوری می‌شود؛ در آخر، از تلفیق اطلاعات استخراج‌شده از گزارش‌های خبری مرتبط، اطلاعات رخداد بدست می‌آید.

فرض اول پژوهش بر این است که فارغ از نوع رخداد طبیعی، خبرگزاری‌ها تنها به رخدادهایی می‌پردازند که برای مخاطب مهم باشد. از این رو ممکن است تحت تاثیر یک زلزله شدید، پس‌لرزه‌هایی که در حالت عادی اهمیت پایینی برای گزارش داشتند، با توجه به تمرکز فضای رسانه‌ای به آن منطقه، بطور منظم گزارش شوند. از طرفی، در شرایطی که رخدادهای طبیعی بویژه رخدادهای هواشناسی، اغلب به دنبال هم اتفاق می‌افتند، گزارش‌های خبری معمولاً به رخداد موثرتر یا رخداد نهایی می‌پردازند. لذا، نتایج این تحقیق نمی‌تواند بیانگر تمام رخدادهای طبیعی باشد.

با توجه به حجم گزارش‌های خبری و نیز نحوه اعلام وقوع رخدادها، تنها رخدادهای آتش‌سوزی (محدود به آتش‌سوزی منابع طبیعی و مزارع کشاورزی)، آذرخش، بارش رگباری، تگرگ، ریزش بهمن، زمین‌لرزه، سیل، تندباد، کولاک، گردوغبار و یخبندان برای استخراج رویدادها در نظر گرفته شد و رخدادهای فرونشست، زمین‌لغزش و خشکسالی کنار گذاشته شد.



شکل ۱- روندنمای چارچوب طراحی‌شده برای استخراج رخدادهای طبیعی از وبگاه‌های خبری

در این پژوهش، اگرچه در سطح استخراج رخدادها، جاینام‌های شهرستان به عنوان محل رخداد لحاظ شده‌اند ولی نتایج نهایی تنها در سطح استان ارائه شده است. دلیل این امر، نخست به نحوه رفتار پدیده‌های طبیعی از نظر وسعت تحت پوشش برمی‌گردد. به بیانی دیگر برخی رخدادها به ازای هر شهرستان تعریف شده و گروهی دیگر می‌توانند چند شهرستان را شامل شوند. همچنین، نبود چارچوب و رویه ثابتی در ارائه اطلاعات رخدادها (صرف نظر از اشتباهات نگارشی و علمی) باعث می‌شود تا علیرغم عنوان شدن نام شهرها در بسیاری از گزارش‌ها، در بعضی موارد به ذکر نام استان اکتفا شود.

^۱ Toponym

۳-۱- تهیه داده

برای این پژوهش، وبگاه خبرگزاری مهر به علت پوشش خبری مناسب از لحاظ مکانی و زمانی، به عنوان منبع اصلی انتخاب شد و از گزارش‌های خبری خبرگزاری ایسنا برای آموزش مدل‌های مبتنی بر یادگیری ماشین مرحله طبقه‌بندی متن استفاده شد. در وبگاه هر دو خبرگزاری، براساس کلیدواژه‌های مرتبط با هر نوع رخداد طبیعی جستجو بصورت خودکار انجام گرفت و تنها اخباری که از آغاز ۱۳۹۰ تا پایان ۱۳۹۹ منتشر شده بودند، لحاظ شدند. به ازای هر گزارش خبری، عنوان خبر، متن خبر، تاریخ انتشار، واحد استانی منتشرکننده و نشانی اینترنتی خبر ذخیره شد. تعداد گزارش‌های خبری گردآوری شده به ازای هر نوع رخداد در جدول ۱ ارائه شده است.

جدول ۱- تعداد گزارش‌های خبری وبکاو شده از دو خبرگزاری مهر و ایسنا

نام خبرگزاری	مهر	ایسنا
آتش‌سوزی	۲۲۴۷	۸۶۵
آذرخش	۲۱۶	۱۱۳
بارش رگباری	۴۴۸	۲۰۳
تگرگ	۳۹۴	۱۸۴
ریزش بهمن	۲۸۶۹	۹۳۲
زمین لرزه	۷۱۹۶	۲۹۶۴
سیل	۶۳۳۷	۴۰۵۴
تندباد	۱۱۵۶	۶۳۸
کولاک	۶۴۶	۴۱۱
گردوغبار	۱۶۴	۸۸
یخبندان	۲۶۷	۱۳۰

۳-۲- پاکسازی داده

با توجه به جستجوی مبتنی بر کلیدواژه، موارد ذخیره شده شامل گزارش‌های خبری نامرتب بود. اخبار شامل کلیدواژه‌ها ولی در معنایی متفاوت، گزارش‌های مربوط به مانورهای آمادگی در مقابل رخدادها، موارد شامل تنها نقل قول افراد و اغلب تبیین ضرورت‌ها، تکذیب شایعات، انتخاب شهرها یا استان‌های معین، بررسی‌های مربوط به تاب‌آوری مناطق مسکونی و گردهمایی‌ها از جمله موارد غیرمرتبط بودند که در مرحله پاکسازی داده، با بررسی اولیه خودکار، بخش قابل توجهی از این موارد حذف شد و در آخر، موارد باقیمانده مورد بازبینی قرار گرفت.

۳-۳- استانداردسازی متن اخبار

در این مرحله، استانداردسازی‌های لازم بر روی واژگان و عبارات متن اخبار بصورت خودکار انجام گرفت. از جمله می‌توان به تغییر نقطه به ممیز در اعداد اعشاری و تبدیل اعداد انگلیسی و عربی به فارسی، تغییر عبارات اختصاری به شکل کاملشان و تصحیح املای واژه‌های پرکاربرد اشاره کرد. با توجه به اینکه تنها جملات آغازین متن خبر دربرگیرنده اطلاعات مفید می‌باشند، محدودسازی متن اخبار انجام شد و فقط سه جمله ابتدایی خبر برای پردازش‌های بعدی در نظر گرفته شد.

۳-۴- طبقه‌بندی گزارش‌های خبری

گزارش‌های خبری منتشرشده را می‌توان از این نظر که درباره رخدادی جدید هستند یا اطلاعاتی تکمیلی درباره رخدادی جاری ارائه می‌کنند، به دو گروه تقسیم کرد. در نتیجه این طبقه‌بندی، گزارش‌هایی می‌توانند به عنوان اولین گزارش یک رخداد لحاظ شوند که وقوع رخدادی جدید را اعلام کنند. همچنین، بر این اساس که گزارش خبری در قالب چه موضوعی به رخداد طبیعی پرداخته است، می‌توان ۸ گروه را در نظر گرفت: ۱- پیش‌بینی وقوع حادثه؛ ۲- بدون خسارت بودن حادثه؛ ۳- خسارات حادثه؛ ۴- اعلام آمادگی سازمان‌ها برای کمک‌رسانی؛ ۵- کمک‌رسانی توسط سازمان‌ها یا مردم؛ ۶- بازدید توسط مسئولین ذیربط؛ ۷- مصوبات و ۸- بازسازی مناطق حادثه‌دیده. بر مبنای این طبقه‌بندی، در مرحله شناسایی جاینام‌ها و نیز در نامزد شدن گزارش‌های خبری به عنوان اولین خبر یک رخداد، قواعد محدود کننده تعریف می‌شود. آخرین طبقه‌بندی، شناسایی انواع رخدادها طبیعی است که در خبر مطرح شده اند. با توجه به بررسی اولیه‌ای که بر روی چند روش مطرح در طبقه‌بندی متن گزارش‌های خبری این تحقیق انجام گرفت، در مقایسه با ماشین بردار پشتیبان و جنگل‌های تصادفی، روش بیز ساده به دلیل ارائه نتایجی با دقت و بازیابی بالاتر، برای دو طبقه‌بندی اول و دوم انتخاب شد. برای آموزش مدل‌ها، اخبار مربوط به سال‌های ۱۳۹۰ تا ۱۳۹۲ خبرگزاری ایسنا برچسب‌گذاری شد. برای پیاده‌سازی مدل، برداری‌سازی متن اخبار بر اساس ان-گرم‌های^۱ یکتایی، دوتایی و سه‌تایی

^۱ N-gram

انجام شد. نتیجه ارزیابی عملکرد مدل در دو طبقه‌بندی اول و دوم در جدول ۲ ارائه شده است. برای طبقه‌بندی سوم، مدلی قاعده‌محور طراحی شد تا بر پایه کلیدواژه‌های مرتبط (نظیر «لرزانند» یا «لرزید» برای زلزله و «بارش شدید» برای بارش‌های رگباری)، انواع رخداد‌های طبیعی مطرح‌شده در گزارش خبری شناسایی شود.

جدول ۲- نتیجه ارزیابی عملکرد مدل‌های طبقه‌بندی متن اخبار براساس دو معیار نوع رخداد گزارش‌شده و نوع موضوع مطرح‌شده

معیار طبقه‌بندی گزارش خبری	درباره رخداد جدید یا جاری	موضوع مطرح‌شده
بازیابی	۰/۸۶۹	۰/۷۲۵
دقت	۰/۸۸۷	۰/۷۴۸
F-score	۰/۸۷۵	۰/۷۱۶

۳-۵- شناسایی جاینام‌ها

با توجه به این که سطح مدیریتی تعریف‌شده برای محل رخدادها، شهرستان بود، تنها نام جغرافیایی استان‌ها، شهرستان‌ها و شهرها از متن اخبار استخراج شد. برای این منظور، مجموعه این جاینام‌ها در ساختاری سلسله‌مراتبی از وبگاه وزارت کشور دانلود شد. مدل قاعده‌محور طراحی شده پس از شناسایی اولیه جاینام‌ها، رفع ابهام را براساس موضوع گزارش خبری و واحد استانی منتشرکننده خبر اعمال کرد.

۳-۶- پیش‌پردازش استخراج رخداد‌های طبیعی

در این مرحله، پیش‌پردازش‌های لازم در سطح جمله انجام شد تا موارد اضافی حذف شده و ساختار داده مناسب برای استخراج رخدادها فراهم شود. از جمله موارد اضافی می‌توان به اعلام وضعیت دمای فعلی مرکز استان یا سردترین و گرم‌ترین شهر استان پس از اعلام وقوع رخدادی مثل تگرگ در شهری خاص، ارائه اطلاعاتی درباره رخدادی مهم در سال گذشته پس از اعلام وقوع رخدادی جدید یا گزارش سالیانه درباره انواع رخدادها اشاره کرد. در مرحله ساختاردهی، در گام نخست، از بین جاینام‌های شناسایی‌شده کل گزارش خبری پس از نگاشت نام شهرها به نام شهرستان مربوطه، جاینام‌های مرتبط با هر رخداد طبیعی مشخص شد (استخراج «نوع رخداد-جاینام»‌ها). در گام دوم، از بین

جاینام‌های مرتبط، به کمک عبارات با قاعده^۱، جاینام‌ها در چهار گروه دسته‌بندی شدند. گروه اول تحت عنوان جاینام‌های اصلی، جاینام‌هایی هستند که حادثه در آنجا رخ داده یا احساس شده است به طوری که باعث خسارت شده باشد. گروه دوم جاینام‌هایی هستند که مخاطره طبیعی در آنجا به شکل ضعیف‌تری احساس شده باشد. گروه سوم شامل جاینام‌هایی می‌شود که به محل حادثه کمک‌رسانی کرده‌اند. جاینام‌هایی که در هیچ‌کدام از این سه گروه شناسایی نشدند، در گروه چهارم قرار گرفتند. در آخر، به ازای هر مورد «نوع رخداد-جاینام»، اطلاعات بیان‌شده در متن خبر با استفاده از عبارات با قاعده استخراج شد (جدول ۳). در صورت استخراج تاریخ و زمان وقوع حادثه از متن خبر، این موارد جایگزین تاریخ و زمان انتشار خبر شدند.

جدول ۳- موارد تعریف‌شده برای استخراج از متن خبر به ازای هر نوع رخداد طبیعی

نوع رخداد طبیعی	موارد تعریف‌شده برای استخراج از متن خبر
آتش‌سوزی	تاریخ و زمان حادثه، تلفات جانی حادثه، تعداد زیرحادثه‌ها، خسارات مالی حادثه، نوع محدوده مورد آتش‌سوزی، سطح خسارت‌دیده
آذرخش	تاریخ و زمان حادثه، تلفات جانی حادثه
بارش رگباری	تاریخ و زمان حادثه، تلفات جانی حادثه، خسارات مالی حادثه، سطح خسارت‌دیده
تگرگ	تاریخ و زمان حادثه، تلفات جانی حادثه، خسارات مالی حادثه، سطح خسارت‌دیده
ریزش بهمن	تاریخ و زمان حادثه، تلفات جانی حادثه، تعداد زیرحادثه‌ها
زمین‌لرزه	تاریخ و زمان حادثه، تلفات جانی حادثه، عمق کانونی زلزله، بزرگی زلزله، موقعیت جغرافیایی کانون زلزله
سیل	تاریخ و زمان حادثه، تلفات جانی حادثه، خسارات مالی حادثه، سطح خسارت‌دیده
تندباد	تاریخ و زمان حادثه، تلفات جانی حادثه، خسارات مالی حادثه، سطح خسارت‌دیده، سرعت باد
کولاکی	تاریخ و زمان حادثه، تلفات جانی حادثه، خسارات مالی حادثه، سطح خسارت‌دیده
گردوغبار	تاریخ و زمان حادثه، تلفات جانی حادثه
یخبندان	تاریخ و زمان حادثه، تلفات جانی حادثه

^۱ Regular expressions

۳-۷- استخراج رخدادهای طبیعی

به ازای هر نوع رخداد طبیعی، ابتدا بر اساس نوع گزارش خبری و موضوع خبر، رخداد جدید شناسایی شده و بر مبنای پارامترهای زمانی تعریف شده برای هر نوع رخداد طبیعی شامل فاصله زمانی هر گزارش از گزارش اول و آخرین گزارش گردآوری شده، گزارش‌های بعدی به نخستین گزارش رخداد اضافه شدند. از نقطه نظر مکانی، برای رخدادهای زمین‌لرزه، تندباد، کولاک و گردوغبار امکان اتخاذ چندین شهرستان به عنوان مکان حادثه تعریف شد و برای سایر انواع رخداد، برای هر جابجای اصلی شناسایی شده، یک رخداد مستقل در نظر گرفته شد. برای هر رخداد زمین‌لرزه، گزارش‌های خبری مربوط به پس‌لرزه‌ها با مشخص نمودن وقوع پس‌لرزه به مجموعه اخبار حادثه اصلی اضافه شد. در نهایت، اطلاعات هر رخداد بر مبنای اطلاعات استخراج شده از گزارش‌های خبری گردآوری شد.

۴- نتایج تحقیق

در این بخش، نتایج حاصل از پیاده‌سازی چارچوب توسعه داده شده برای استخراج رخدادهای طبیعی ایران از اخبار ارائه شده است. نخست، تعداد رخدادهای گزارش شده با موارد ثبت شده رسمی مقایسه شده است. در بخش دوم، موضوعات مطرح شده در گزارش‌های خبری به تفکیک نوع رخداد ارائه شده است. در بخش سوم، توزیع زمانی رخدادهای بر اساس ماه و سال مورد بررسی قرار گرفته است. تعداد رخدادهای طبیعی به تفکیک نوع رخداد برای هر استان و نیز سهم هر استان از انواع رخداد در بخش چهارم ارائه شده است. در ادامه، هم‌رخدادی حوادث طبیعی مورد بحث قرار گرفته است. در آخر، اطلاعات استخراج شده درباره رخدادهای زمین‌لرزه در قالب نمودارهای پراکندگی زمانی نمایش داده شده اند. شکل ۲ دو نمونه از رخدادهای استخراج شده از نوع زمین‌لرزه و آتش‌سوزی را نشان می‌دهد.

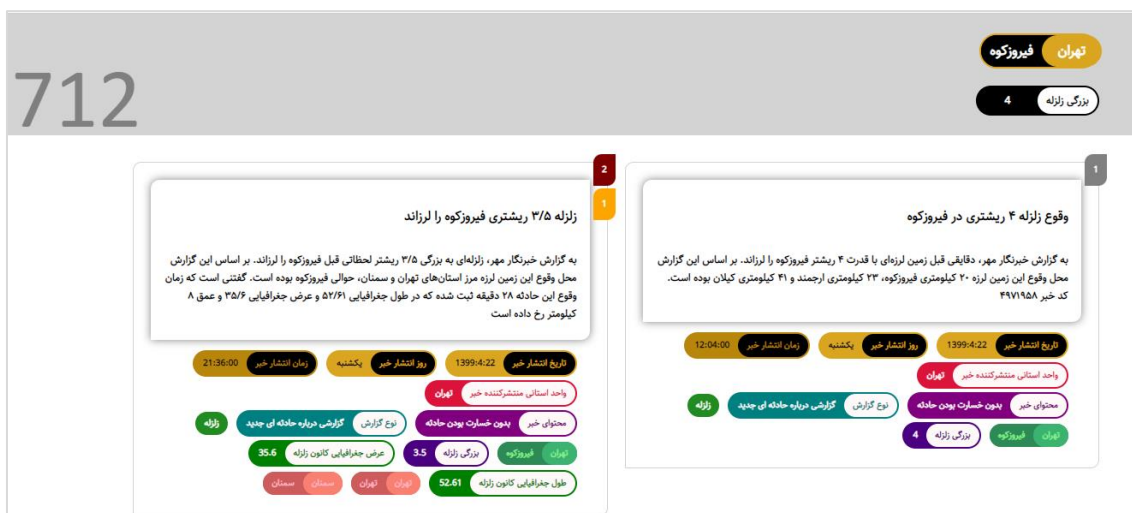
برای ارزیابی مجموعه رویدادهای استخراج شده، از نقشه‌های موضوعی شش ماه نخست سال ۱۳۹۹ سازمان مدیریت بحران کشور استفاده شده است. در این نقشه‌ها، بازه‌های مختلفی برای تعداد رخدادهای تعریف شده است و به ازای هر بازه، وضعیت استان‌های مربوطه با رنگ مشخصی نمایش داده شده است. از بین رخدادهای مورد مطالعه در

این پژوهش، تنها رخدادهای تندباد، تگرگ، یخبندان، سیل و زمین‌لرزه مورد مقایسه قرار گرفته است. برای این منظور، تعداد رخدادهای استخراج شده در همان بازه‌های تعریف شده در نقشه‌ها مشخص شده و براساس اختلاف طبقه‌ها، بررسی انجام گرفته است. دلایل اختلاف در تعداد رخدادهای در ادامه بیان می‌شود.

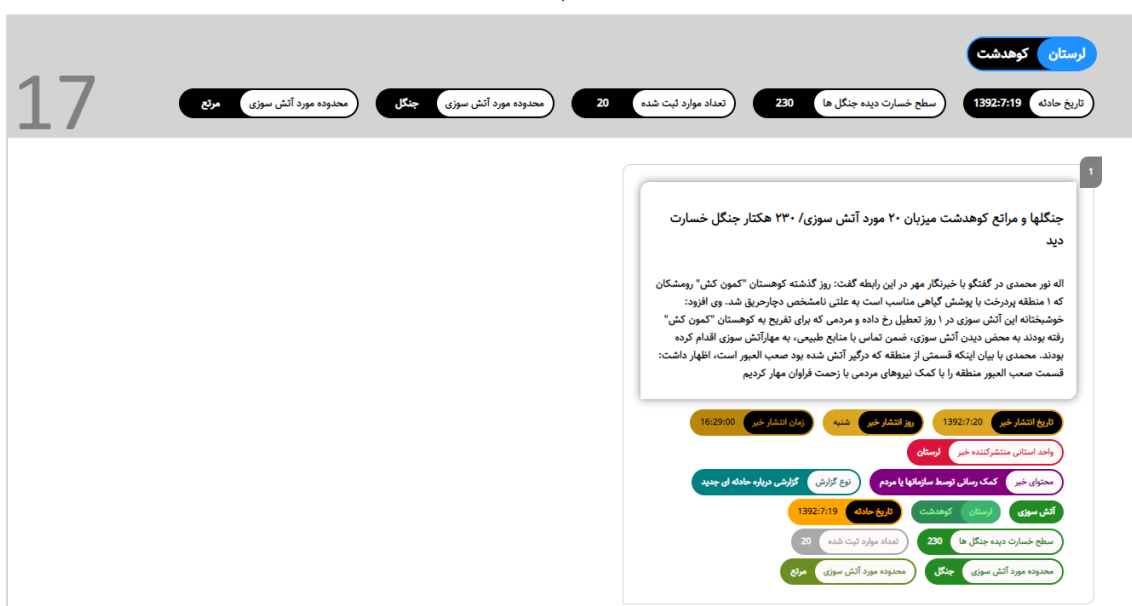
با توجه این که اخبار هر استان توسط واحد خبرگزاری آن استان تحت پوشش خبری قرار می‌گیرد، فارغ از میزان فعال بودن واحد مربوطه، اولویت‌های ارائه گزارش حتی برای اخبار مربوط به حوادث طبیعی نیز ممکن است وابسته به استان تغییر کند. از این رو ممکن است برای نوع رخداد غالب در یک استان، فرکانس گزارش خبر در مقایسه با نوع رخدادی که از نظر خبری تازگی دارد، پایین باشد. از طرفی، ممکن است شهرستان‌ها توسط خبرگزاری واحد استانی در اولویت‌بندی خاصی قرار بگیرند. بنابراین، میزان توجه واحد استانی به رخدادهای طبیعی صرف نظر از جوانب سیاسی و مدیریتی را می‌توان اولین دلیل اختلاف تعداد رخدادهای عنوان کرد.

دلیل دوم تفاوت در تعیین عمر رخدادهای و در نتیجه، نحوه شمارش رخدادهای است. برای مثال، رخداد تندباد ممکن است توسط چارچوب طراحی شده شناسایی شود ولی براساس پارامترهای زمانی تعریف شده، گزارش‌های مربوط به رخدادهای دو یا سه روز بعد که توسط سازمان‌های ذیربط ثبت می‌شود، توسط چارچوب برای رخداد اول اختصاص یابد. این مسئله به خاطر این اتفاق می‌افتد که اغلب آغاز رخدادهای و در مواردی ادامه رخدادهای در اخبار اعلام می‌شود ولی پایان رخدادهای بجز از حوادثی مثل آتش‌سوزی، مشخص نمی‌شود.

در حالت کلی، صرف نظر از پیش‌بینی رخدادهای هواشناسی، خبرگزاری‌ها اغلب وقوع حوادثی را اعلام می‌کنند که اثری بر منابع و دارایی‌های شخصی یا عمومی داشته است. از این رو ممکن است چندین زلزله بالای سه ریشتر در یک استان اتفاق بیافتد ولی تنها برخی موارد ارزش خبری پیدا کنند.



الف)



ب)

شکل ۲- دو نمونه از رخدادهای طبیعی استخراج شده از گزارش های خبری. الف) نمونه ای از یک رخداد زمین لرزه و یک پس لرزه و ب) نمونه ای از یک رخداد آتش سوزی

جدول ۴- مقایسه تعداد رخداد های استخراج شده از اخبار با تعداد رخداد های اعلام شده توسط سازمان مدیریت بحران کشور برای شش ماه نخست ۱۳۹۹. به ازای هر استان، ردیف نخست شامل بازه های اعلام شده توسط سازمان و ردیف دوم شامل بازه های تعداد رخداد های استخراج شده است. رنگ های خاکستری، زرد، نارنجی، قرمز و بنفش به ترتیب بیانگر این است که براساس طبقات تعریف شده، دو بازه به اندازه صفر، یک، دو، سه و چهار طبقه فاصله دارند. رنگ سبز بیانگر این است که تعداد رخداد های استخراج شده از تعداد رخداد های اعلام شده توسط سازمان بیشتر است

نام استان	تندباد	تگرگ	یخبندان	سیل	زمین لرزه
آذربایجان شرقی	۰-۲	۰	۰-۱	۵-۱۱	۴-۱۳
	۰-۲	۰	۰-۱	۰-۴	۰-۳
آذربایجان غربی	۰-۲	۱-۲	۰-۱	۵-۱۱	۱۴-۳۷
	۰-۲	۱-۲	۰-۱	۰-۴	۰-۳
اردبیل	۹-۱۳	۰	۱۲-۲۸	۲۰-۳۰	۴-۱۳
	۰-۲	۰	۰-۱	۰-۴	۰-۳
اصفهان	۰-۲	۱-۲	۰-۱	۰-۴	۴-۱۳
	۰-۲	۰	۰-۱	۰-۴	۰-۳

۰-۳	۰-۴	۰-۱	۰	۰-۲	البرز
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۴-۱۳	۰-۴	۰-۱	۰	۰-۲	ایلام
۴-۱۳	۰-۴	۰-۱	۰	۰-۲	
۱۴-۳۷	۰-۴	۰-۱	۳-۵	۱۴-۲۲	بوشهر
۰-۳	۰-۴	۰-۱	۱-۲	۰-۲	
۴-۱۳	۰-۴	۰-۱	۰	۰-۲	تهران
۴-۱۳	۰-۴	۰-۱	۱-۲	۰-۲	
۴-۱۳	۵-۱۱	۲-۳	۶-۹	۰-۲	چهارمحال و بختیاری
۰-۳	۰-۴	۰-۱	۱-۲	۰-۲	
۱۴-۳۷	۱۲-۱۹	۰-۱	۰	۰-۲	خراسان جنوبی
۴-۱۳	۰-۴	۰-۱	۰	۰-۲	
۱۴-۳۷	۳۱-۵۶	۱۲-۲۸	۱۰-۲۰	۳-۸	خراسان رضوی
۴-۱۳	۰-۴	۰-۱	۱-۲	۰-۲	
۱۴-۳۷	۱۲-۱۹	۰-۱	۶-۹	۰-۲	خراسان شمالی
۴-۱۳	۰-۴	۰-۱	۱-۲	۰-۲	
۱۴-۳۷	۱۲-۱۹	۰-۱	۰	۳-۸	خوزستان
۴-۱۳	۰-۴	۰-۱	۰	۰-۲	
۰-۳	۲۰-۳۰	۴-۱۰	۶-۹	۳-۸	زنجان
۰-۳	۰-۴	۰-۱	۱-۲	۰-۲	
۱۴-۳۷	۱۲-۱۹	۱۱	۳-۵	۰-۲	سمنان
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۰-۳	۱۲-۱۹	۰-۱	۰	۳-۸	سیستان و بلوچستان
۰-۳	۰-۴	۰-۱	۰	۳-۸	
۷۷-۱۳۳	۱۲-۱۹	۰-۱	۰	۰-۲	فارس
۱۴-۳۷	۰-۴	۰-۱	۰	۰-۲	
۴-۱۳	۵-۱۱	۰-۱	۰	۰-۲	قزوین
۴-۱۳	۰-۴	۰-۱	۰	۰-۲	
۴-۱۳	۵-۱۱	۰-۱	۱-۲	۳-۸	قم
۴-۱۳	۰-۴	۰-۱	۰	۰-۲	
۰-۳	۵-۱۱	۰-۱	۱-۲	۰-۲	کردستان
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۳۸-۷۶	۲۰-۳۰	۰-۱	۱-۲	۱۴-۲۲	کرمان
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۱۴-۳۷	۰-۴	۰-۱	۰	۰-۲	کرمانشاه
۴-۱۳	۰-۴	۰-۱	۱-۲	۰-۲	
۱۴-۳۷	۲۰-۳۰	۰-۱	۰	۰-۲	کهگیلویه و بویراحمد
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۱۴-۳۷	۵-۱۱	۰-۱	۰	۰-۲	گلستان
۴-۱۳	۰-۴	۰-۱	۱-۲	۰-۲	
۱۴-۳۷	۰-۴	۰-۱	۰	۰-۲	گیلان
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۴-۱۳	۰-۴	۰-۱	۰	۰-۲	لرستان
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۴-۱۳	۵-۱۱	۰-۱	۰	۳-۸	مازندران

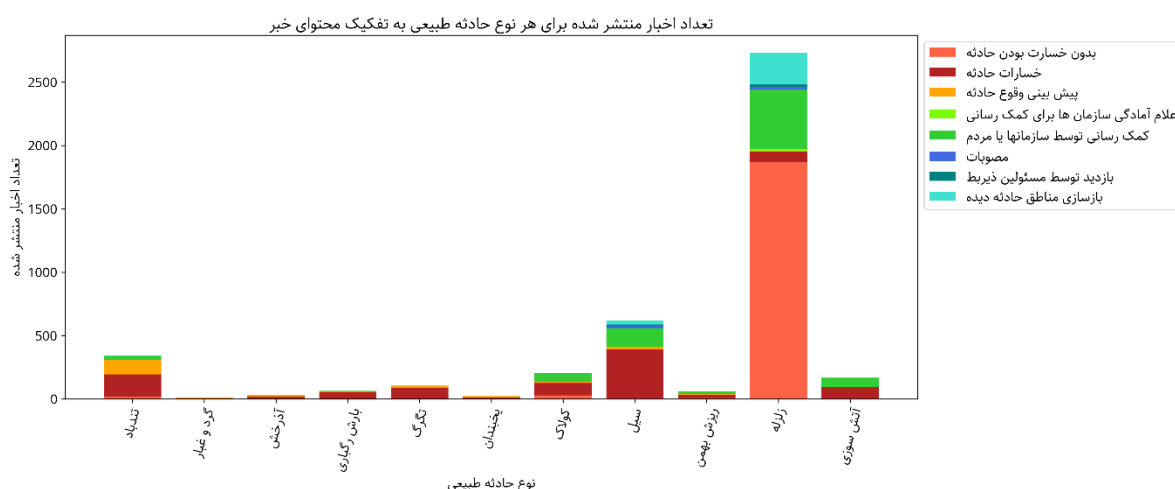
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۰-۳	۰-۴	۰-۱	۱-۲	۲۳-۲۶۱	مرکزی
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۳۸-۷۶	۰-۴	۰-۱	۰	۰-۲	هرمزگان
۴-۱۳	۰-۴	۰-۱	۰	۰-۲	
۰-۳	۰-۴	۰-۱	۱-۲	۳-۸	همدان
۰-۳	۰-۴	۰-۱	۰	۰-۲	
۴-۱۳	۵-۱۱	۰-۱	۱-۲	۱۴-۲۲	یزد
۰-۳	۰-۴	۰-۱	۱-۲	۰-۲	

استخراج شده از موارد ثبت شده است. در حالی که طبق گزارش سازمان رخداد تگرگی برای استان های تهران، کرمانشاه و گلستان ثبت نشده است، رخدادهایی از گزارش های خبری استخراج شده است.

۱-۴- موضوعات مطرح شده در گزارش های خبری

در شکل ۳، فراوانی اخبار منتشر شده هر نوع رویداد به تفکیک موضوعات پرداخته شده در خبر نشان داده شده است. بر طبق این شکل، تعداد اخبار منتشر شده درباره رخدادهای زمین لرزه تفاوت چشمگیری با سایر رخدادهای دارد که بخش قابل توجهی از این اخبار نیز دربرگیرنده اطلاعاتی مبنی بر بدون خسارت بودن حادثه بوده اند. برای بیشتر نوع رخدادهای، گزارش خبری زمانی منتشر شده است که حادثه خساراتی را در پی داشته است یا به خاطر اهمیت مسئله پیش بینی وقوع حادثه اعلام شده است.

مطابق جدول ۴، با توجه به بازه زمانی شش ماهه، اغلب بازه ها برای رخدادهای هواشناسی به طبقه نخست اشاره دارد. از این رو، بازه های مربوط به استخراج رخدادهای در بیشتر موارد همان بازه اعلام شده توسط سازمان است. البته برای طبقه های بعدی نیز موارد عدم اختلاف هست که از آن جمله می توان به «یلام-زمین لرزه»، «تهران-زمین لرزه» و «سیستان و بلوچستان-تندباد» اشاره کرد. تفاوت یک طبقه بین بازه ها بیشترین فراوانی را در بین موارد اختلافی دارد که در بین انواع رخداد، زمین لرزه و در بین استان ها، استان های مازندران و قم بیشترین موارد مربوط به این سطح از اختلاف را دارند. در سطوح بعدی تفاوت بازه ها، رخداد سیل بیشترین موارد را شامل می شود. بخشی از حجم بالای اختلاف مربوط به این رخداد را شاید بتوان در تجمیع آمار سیل و آبگرفتگی در موارد اعلام شده دانست در حالی که در این پژوهش، آبگرفتگی معابر شهری جزء رخدادهای مورد مطالعه نبوده است. نکته قابل توجه پیشی گرفتن تعداد رخدادهای

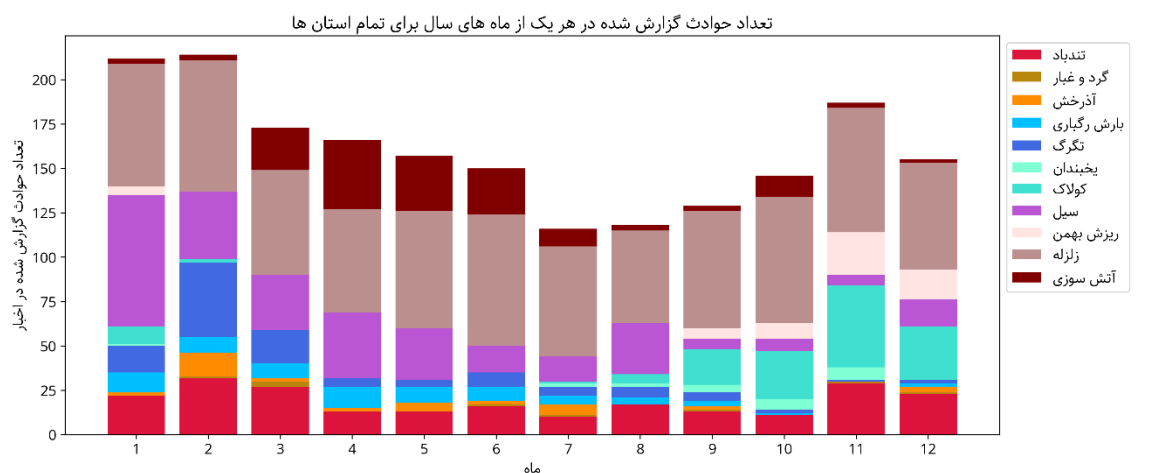


شکل ۳- تعداد اخبار منتشر شده برای هر نوع حادثه طبیعی به تفکیک محتوای خبر

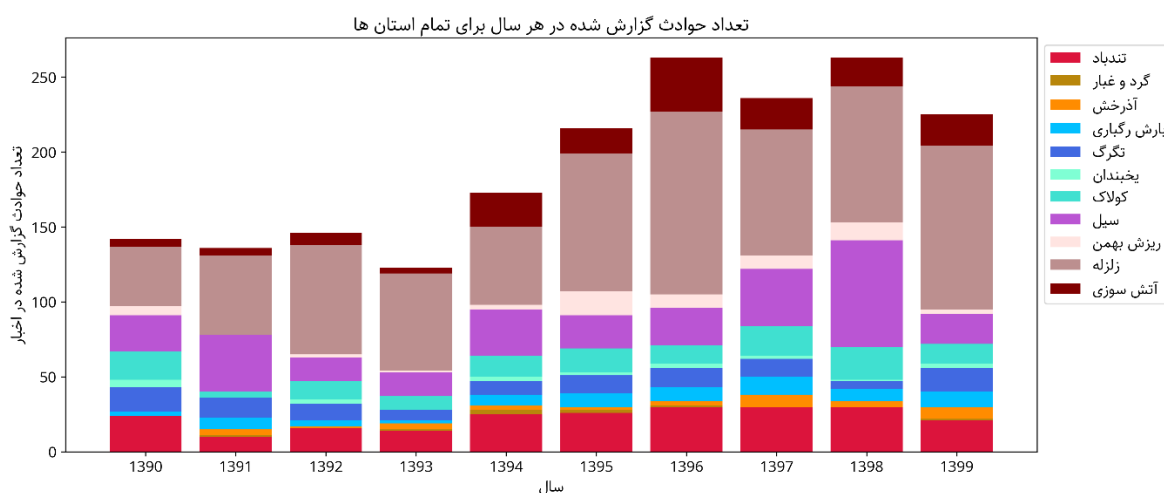
۴-۲- توزیع زمانی رخدادها

توزیع زمانی رخدادهای طبیعی کل کشور براساس ماه و سال به ترتیب در شکل‌های ۴ و ۵ نشان داده شده است. با توجه به شکل ۴، بصورت کلی می‌توان گفت که بیشینه تعداد رخدادها در ماه‌های آغازین و پایانی سال و کمینه آن‌ها در ماه‌های میانی سال بوده است. رخدادهای هواشناسی با توجه به تاثیری که از شرایط هوا دارند، به شدت به تغییرات فصل وابسته اند. در مقابل، تعداد رخداد زمین‌لرزه برای تمامی ماه‌ها ثابت است. برای نوع رخداد آتش‌سوزی، ماه‌های تیر، مرداد، شهریور و خرداد به ترتیب بیشترین موارد وقوع حادثه را داشته اند. بیشترین میزان وقوع رخداد برای سیل، تگرگ،

بارش رگباری و کولاک به ترتیب در ماه‌های فروردین، اردیبهشت، تیر و بهمن اتفاق افتاده است. همچنین برای نوع رخداد یخبندان و ریزش بهمن نیز بیشترین موارد گزارش‌شده مربوط به ماه بهمن بوده است. در توزیع زمانی بر اساس سال، شکل ۵ نشان‌دهنده روندی افزایشی است. در این میان، افزایش پوشش خبری در طی این سال‌ها را نیز باید در نظر داشت. با این حال، برای برخی از انواع رخداد این مسئله تاثیری نداشته است. بر اساس این نمودار می‌توان وقوع یک نوع رخداد را در سطح گسترده‌ای استنباط نمود. از جمله می‌توان به تاثیر چشمگیر سیل‌های سال ۱۳۹۸ که ۲۰ استان از کشور را درگیر کرده بود، اشاره کرد.



شکل ۴- تعداد حوادث گزارش شده در هر یک از ماه‌های سال برای تمام استان‌ها



شکل ۵- تعداد حوادث گزارش شده در هر سال برای تمام استان‌ها

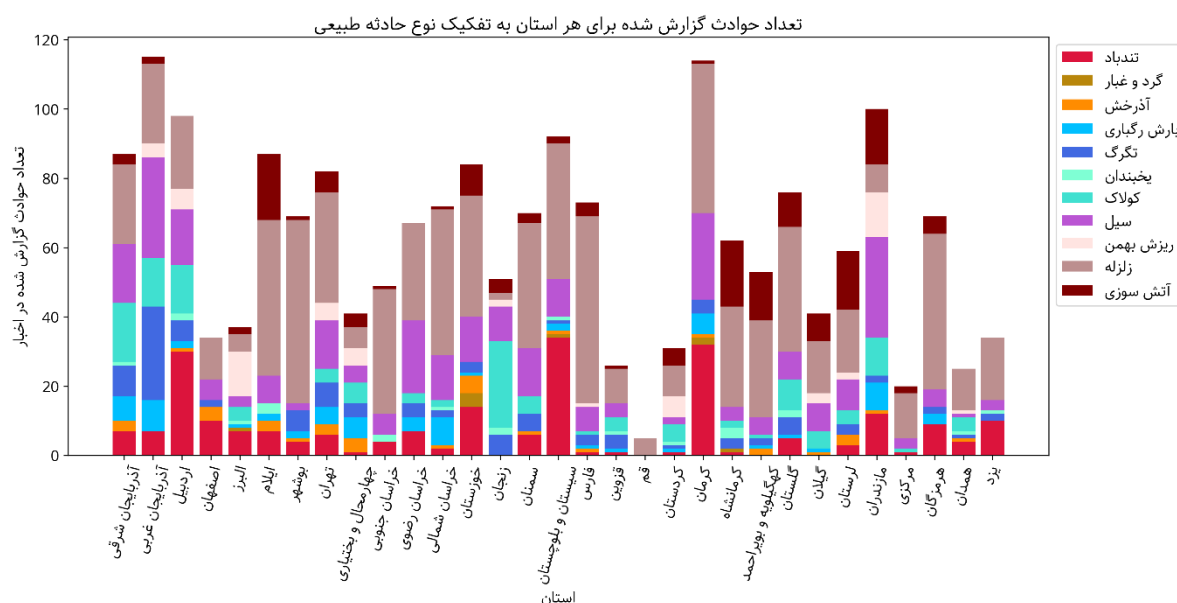
۴-۳- توزیع مکانی رخدادها

شکل‌های ۶ و ۷ نشان‌دهنده توزیع مکانی رخدادها در سطح مدیریتی استان هستند. در شکل ۶، به ازای هر

استان، تعداد موارد اتفاق افتاده هر نوع رخداد ارائه شده است. دو استان آذربایجان غربی و کرمان بیشترین و استان قم کمترین رخداد طبیعی گزارش‌شده را داشته اند.

در شکل ۷، سهم هر استان از کل رخداد‌های گزارش شده هر نوع حادثه طبیعی بصورت نقشه موضوعی نشان داده شده است. در نوع رخداد تندباد، استان‌های سیستان و بلوچستان، کرمان و اردبیل بیشترین وقوع حادثه را داشته اند. بیشترین موارد آتش‌سوزی مربوط به استان‌های ایلام، کرمانشاه، لرستان و مازندران بوده است. بیشترین وقوع زمین‌لرزه برای استان‌های فارس، بوشهر، کرمان، هرمزگان،

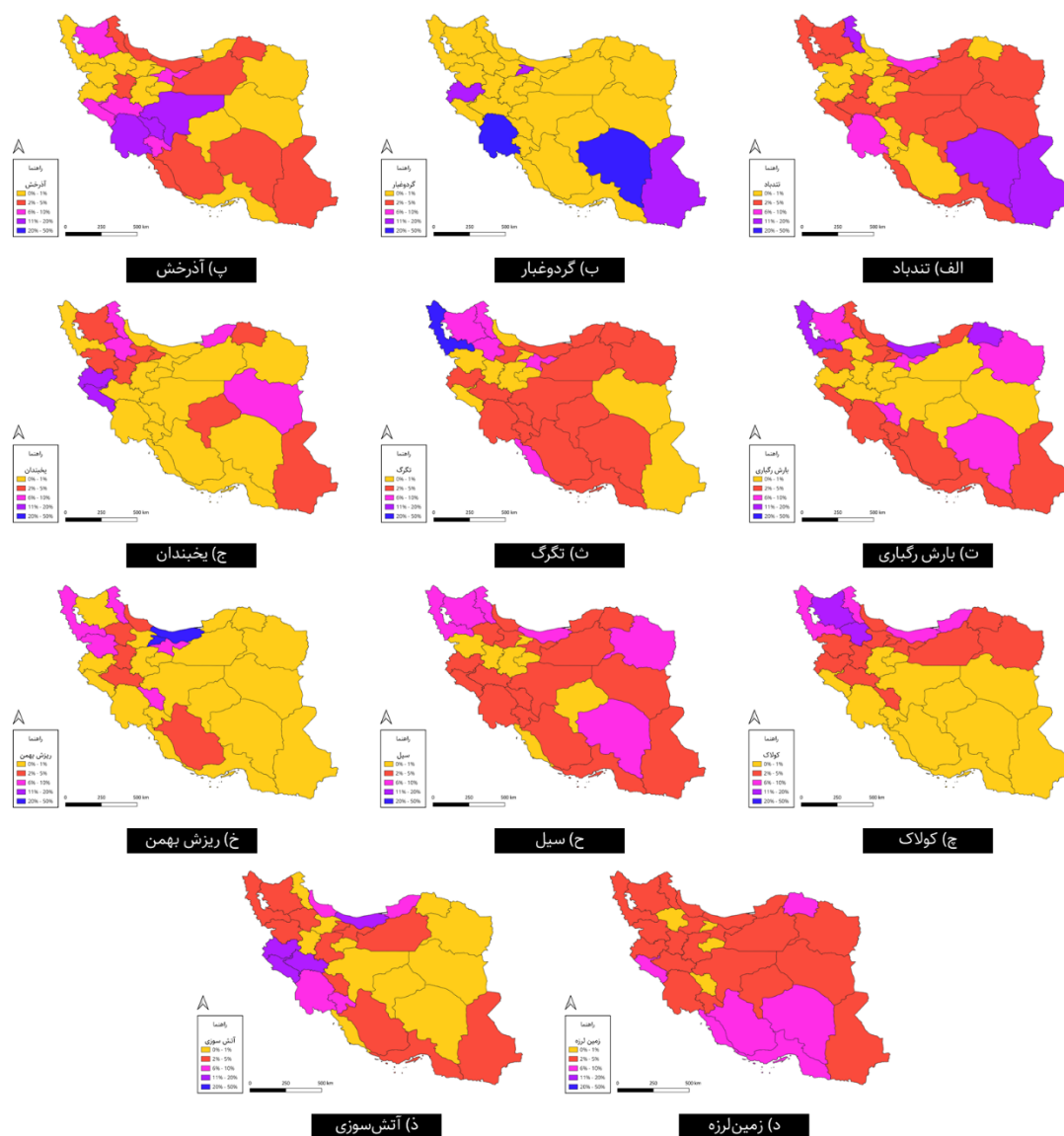
خراسان شمالی و ایلام گزارش شده است. استان‌های آذربایجان غربی، مازندران و کرمان بیشترین موارد وقوع سیل را داشته اند. براساس گزارش‌های خبری، دو استان آذربایجان شرقی و زنجان بیشینه موارد وقوع کولاک و دو استان مازندران و البرز بالاترین میزان رخداد ریزش بهمن را داشته اند.



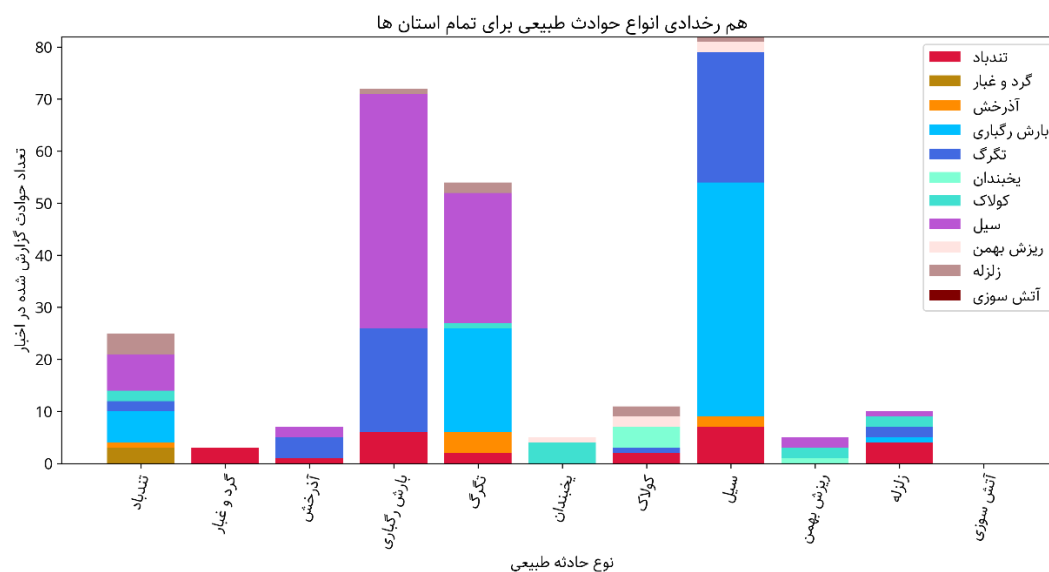
شکل ۶- تعداد حوادث گزارش شده برای هر استان به تفکیک نوع حادثه طبیعی

۴-۴- هم‌رخدادی حوادث طبیعی

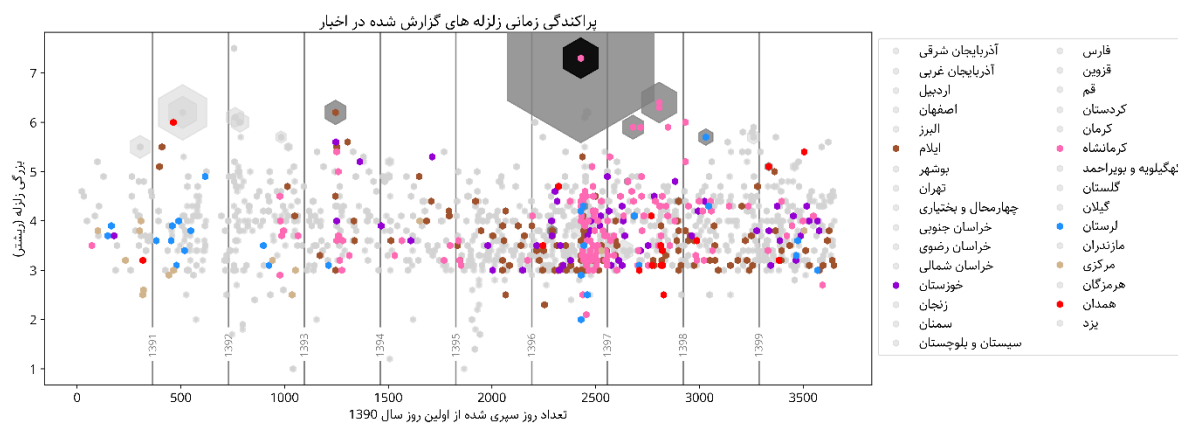
بر این اساس که در یک روز چه نوع رخداد‌هایی در سطح استان اتفاق افتاده است، هم‌رخدادی حوادث طبیعی برآورد شده است. مطابق شکل ۸، آتش‌سوزی بدون هم‌رخدادی و دو نوع رخداد گرد و غبار و یخبندان کمترین هم‌رخدادی را با سایر رخدادها داشته اند. هم‌رخدادی بالای سیل، بارش رگباری و تگرگ در مقایسه با دیگر رخدادها چشمگیر است که بر این اساس، ضروری است آمادگی‌های لازم برای مقابله با هر سه رخداد در صورت پیش‌بینی حتی یک مورد از آنها در سطح استانی فراهم شود. بالاترین تنوع از نظر انواع حوادث هم‌رخداد، برای تندباد است اما میزان هم‌رخدادی با هر کدام از انواع دیگر رخدادها قابل توجه نیست.



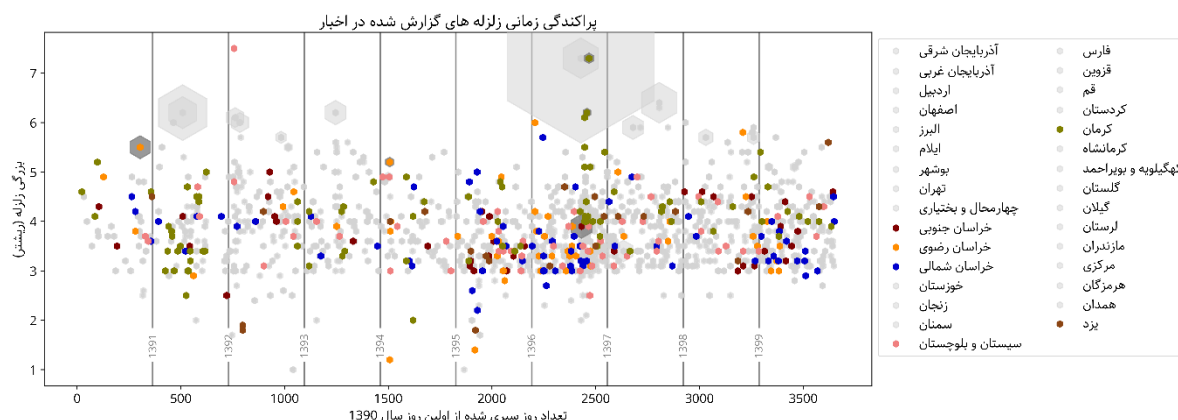
شکل ۷- نقشه‌های موضوعی انواع رخداد طبیعی بیانگر سهم هر استان از کل حوادث گزارش شده کشور برای نوع رخداد مشخص شده



شکل ۸- هم‌رخدادی انواع حوادث طبیعی برای تمام استان‌ها



شکل ۱۲- پراکندگی زمانی زلزله های گزارش شده در اخبار - منطقه ۴



شکل ۱۳- پراکندگی زمانی زلزله های گزارش شده در اخبار - منطقه ۵

۵- نتیجه گیری

این پژوهش، رخدادهای طبیعی ایران را که توسط خبرگزاری مهر از سال ۱۳۹۰ تا ۱۴۰۰ گزارش شده بود را با کمک چارچوب طراحی شده برای استخراج خودکار رخدادها، مورد بررسی قرار داده است. براساس اطلاعات استخراج شده از گزارش های خبری، تحلیل های زمانی و مکانی رخدادها و هم رخدادی آن ها انجام شده است. همچنین، اطلاعات رخدادهای زمین لرزه به عنوان نوع رخدادی که بیشترین گزارش های خبری درباره آن منتشر می شود با مشخص نمودن پراکندگی زمانی، بزرگی و میزان تلفات جانی ارائه شده است.

نتایج کلی پژوهش عبارت اند از:

- مقایسه تعداد رخدادهای ثبت شده توسط سازمان مدیریت بحران کشور با موارد استخراج شده از گزارش های خبری نشان دهنده این است که میزان تفاوت این دو مجموعه رخدادها با توجه به هر استان می تواند متغیر باشد به نحوی که برای برخی استان ها

می توان رخدادهای گزارش شده را برابر رخدادهای اتفاق افتاده دانست ولی برای بعضی استان ها نتایج حاصل از رخدادهای گزارش شده قابل تعمیم نیست.

- بیشترین تعداد کل رخدادها برای دو ماه آغازین و دو ماه پایانی سال است که برای هر نوع رخداد هواشناسی، بازه های زمانی متفاوتی قابل تعریف است. در طی بازه ده ساله، برای بیشتر انواع رخداد بویژه رخداد زمین لرزه، تعداد موارد گزارش شده روندی افزایشی داشته اند که بخشی از آن تحت تاثیر افزایش پوشش خبری است ولی برای برخی رخدادهای هواشناسی تغییرات اقلیمی را می توان عامل موثر دانست.
- براساس تعداد رخدادهای گزارش شده، در تقسیم بندی استان ها به استان های شرقی و غربی، بیشترین تعداد رخداد تنبید برای استان های شرقی گزارش شده است. از طرفی، تنوع سایر رخدادهای هواشناسی بغیر از گردوغبار، برای استان های شمالی در مقایسه با استان های جنوبی کشور بالاتر است. در پدیده

زمین‌لرزه، فارغ از حجم تلفات و خسارات، استان‌های جنوبی بیشترین موارد گزارش‌شده را دارند. هرچند بر اساس بررسی‌های اولیه، موارد گزارش‌شده توسط خبرگزاری مهر درصد بالایی از رخدادهای سایر خبرگزاری‌ها را دربرمی‌گیرد، ولی برای رسیدن به بیشینه تعداد رخدادهای طبیعی گزارش‌شده، برای پژوهش‌های

بعدی پیشنهاد می‌شود که از اخبار چندین خبرگزاری بویژه وبگاه‌های خبری محلی استفاده شود. همچنین، افزایش پوشش خبری می‌تواند بررسی برخی از انواع رخدادهای طبیعی را که در این تحقیق به دلیل گزارش‌های ناکافی به آن‌ها پرداخته نشد را ممکن سازد و امکان برآورد خسارات بر پایه گزارش‌های کافی فراهم شود.

مراجع

- [1] Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A., and Chakraborty, B. (2018), "A review on application of data mining techniques to combat natural disasters," *Ain Shams Engineering Journal*, Vol. 9, No. 3, PP. 365-378.
- [2] Said, N. et al. (2019), "Natural disasters detection in social media and satellite imagery: a survey," *Multimedia Tools and Applications*, Vol. 78, No. 22, PP. 31267-31302.
- [3] Xiang, W. and Wang, B. (2019), "A survey of event extraction from text," *IEEE Access*, Vol. 7, PP. 173111-173137.
- [4] Liu, X., Guo, H., Lin, Y.-r., Li, Y., and Hou, J. (2018), "Analyzing spatial-temporal distribution of natural hazards in China by mining news sources," *Natural Hazards Review*, Vol. 19, No. 3, P. 04018006.
- [5] Franceschini, R., Rosi, A., Catani, F., and Casagli, N. (2022), "Exploring a landslide inventory created by automated web data mining: the case of Italy," *Landslides*, PP. 1-13.
- [6] Puspita, D. and Pranoto, B. E. (2021), "The attitude of Japanese newspapers in narrating disaster events: Appraisal in critical discourse study," *Studies in English Language and Education*, Vol. 8, No. 2, PP. 796-817.
- [7] Yan, Y. and Bissell, K. (2018), "The sky is falling: Predictors of news coverage of natural disasters worldwide," *Communication Research*, Vol. 45, No. 6, PP. 862-886.
- [8] Houston, J. B., Pfefferbaum, B., and Rosenholtz, C. E. (2012), "Disaster news: Framing and frame changing in coverage of major US natural disasters, 2000–2010," *Journalism & Mass Communication Quarterly*, Vol. 89, No. 4, PP. 606-623.
- [9] Talwar, S., Dhir, A., Kaur, P., Zafar, N., and Alrasheedy, M. (2019), "Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior," *Journal of Retailing and Consumer Services*, Vol. 51, PP. 72-82.
- [10] Du, H. et al. (2019), "Twitter vs news: Concern analysis of the 2018 california wildfire event," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 2: IEEE, PP. 207-212.
- [11] Choi, S. and Bae, B. (2015), "The real-time monitoring system of social big data for disaster management," in *Computer science and its applications: Springer*, PP. 809-815.
- [12] Earle, P. S., Bowden, D., and Guy, M. (2011), "Twitter earthquake detection: earthquake monitoring in a social world," *Annals of geophysics*, Vol. 54, No. 6, PP. 708-715.
- [13] Chuo, Y.-J. (2020), "Earthquake Shake Detecting by Data Mining from Social Network Platforms," *Applied Sciences*, Vol. 10, No. 3, P. 812.
- [14] Song, G. and Huang, D. (2021), "A sentiment-aware contextual model for real-time disaster prediction using Twitter data," *Future Internet*, Vol. 13, No. 7, P. 163.
- [15] Yin, J., Karimi, S., Lampert, A., Cameron, M., Robinson, B., and Power, R. (2015), "Using social media to enhance emergency situation awareness," in *Twenty-fourth international joint conference on artificial intelligence*.
- [16] Yu, M., Huang, Q., Qin, H., Scheele, C., and Yang, C. (2019), "Deep learning for real-time social media text classification for situation awareness—using Hurricanes Sandy, Harvey, and Irma as case studies," *International Journal of Digital Earth*, Vol. 12, No. 11, PP. 1230-1247.
- [17] Wang, Z. and Ye, X. (2018), "Social media analytics for natural disaster management," *International Journal of Geographical Information Science*, Vol. 32, No. 1, PP. 49-72.
- [18] Vongkusolkiet, J. and Huang, Q. (2021), "Situational awareness extraction: a comprehensive review of social media data classification during natural hazards," *Annals of GIS*, Vol. 27, No. 1, PP. 5-28.

- [19] Shan, S., Zhao, F., Wei, Y., and Liu, M. (2019), "Disaster management 2.0: A real-time disaster damage assessment model based on mobile social media data—A case study of Weibo (Chinese Twitter)," *Safety science*, Vol. 115, PP. 393-413.
- [20] Yigitcanlar, T. et al. (2022), "Detecting natural hazard-related disaster impacts with social media analytics: the case of Australian states and territories," *Sustainability*, Vol. 14, No. 2, P. 810.
- [21] Neppalli, V. K., Caragea, C., Squicciarini, A., Tapia, A., and Stehle, S. (2017), "Sentiment analysis during Hurricane Sandy in emergency response," *International journal of disaster risk reduction*, Vol. 21, PP. 213-222.
- [22] Dong, Z. S., Meng, L., Christenson, L., and Fulton, L. (2021), "Social media information sharing for natural disaster response," *Natural hazards*, Vol. 107, No. 3, PP. 2077-2104.
- [23] Pennington, C. V. et al. (2022), "A near-real-time global landslide incident reporting tool demonstrator using social media and artificial intelligence," *International Journal of Disaster Risk Reduction*, P. 103089.
- [24] Zhang, M., Huang, Q., and Liu, H. (2022), "A Multimodal Data Analysis Approach to Social Media during Natural Disasters," *Sustainability*, Vol. 14, No. 9, P. 5536.
- [25] Chachra, G. et al. (2022), "Detecting damaged buildings using real-time crowdsourced images and transfer learning," *Scientific Reports*, Vol. 12, No. 1, PP. 1-12.
- [26] Han, X. and Wang, J. (2019), "Using social media to mine and analyze public sentiment during a disaster: a case study of the 2018 Shouguang city flood in china," *ISPRS International Journal of Geo-Information*, Vol. 8, No. 4, P. 185.
- [27] Shen, S., Huang, J., Cheng, C., Zhang, T., Murzintcev, N., and Gao, P. (2021), "Spatiotemporal Evolution of the Online Social Network after a Natural Disaster," *ISPRS International Journal of Geo-Information*, Vol. 10, No. 11, P. 744.
- [28] Xu, J. and Qiang, Y. (2021), "Analysing Information Diffusion in Natural Hazards using Retweets-a Case Study of 2018 Winter Storm Diego," *Annals of GIS*, PP. 1-15.
- [29] Ye, X. and Wei, X. (2019), "A multi-dimensional analysis of El Niño on Twitter: Spatial, social, temporal, and semantic perspectives," *ISPRS International Journal of Geo-Information*, Vol. 8, No. 10, P. 436.
- [30] Franceschini, R., Rosi, A., Catani, F., and Casagli, N. (2022), "Exploring a landslide inventory created by automated web data mining: The case of Italy," *Landslides*, Vol. 19, No. 4, PP. 841-853.
- [31] Badoux, A., Andres, N., Techel, F., and Hegg, C. (2016), "Natural hazard fatalities in Switzerland from 1946 to 2015," *Natural Hazards and Earth System Sciences*, Vol. 16, No. 12, PP. 2747-2768.
- [32] Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., and Lerner-Lam, A. (2010), "A global landslide catalog for hazard applications: method, results, and limitations," *Natural Hazards*, Vol. 52, No. 3, PP. 561-575.
- [33] Froude, M. J. and Petley, D. N. (2018), "Global fatal landslide occurrence from 2004 to 2016," *Natural Hazards and Earth System Sciences*, Vol. 18, No. 8, PP. 2161-2181.
- [34] Han, X. and Wang, J. (2019), "Earthquake information extraction and comparison from different sources based on web text," *ISPRS International Journal of Geo-Information*, Vol. 8, No. 6, P. 252.
- [35] Zêzere, J. L. et al. (2014), "DISASTER: a GIS database on hydro-geomorphologic disasters in Portugal," *Natural hazards*, Vol. 72, No. 2, PP. 503-532.
- [36] Valenzuela, P., Domínguez-Cuesta, M. J., García, M. A. M., and Jiménez-Sánchez, M. (2017), "A spatio-temporal landslide inventory for the NW of Spain: BAPA database," *Geomorphology*, Vol. 293, PP. 11-23.
- [37] Juang, C. S., Stanley, T. A., and Kirschbaum, D. B. (2019), "Using citizen science to expand the global map of landslides: Introducing the Cooperative Open Online Landslide Repository (COOLR)," *PloS one*, Vol. 14, No. 7, P. e0218657.
- [38] de Bruijn, J. A., de Moel, H., Jongman, B., de Ruiter, M. C., Wagemaker, J., and Aerts, J. C. (2019), "A global database of historic and real-time flood events based on social media," *Scientific data*, Vol. 6, No. 1, PP. 1-12.